Quantifying data quality in a citizen science monitoring program: False negatives, false positives and occupancy trends

Cruickshank, Sam S ; Bühler, C ; Schmidt, Benedikt

Abstract: Data collected by volunteers are an important source of information used in species management decisions, yet concerns are often raised over the quality of such data. Two major forms of error exist in occupancy datasets; failing to observe a species when present (imperfect detection—also known as false negatives), and falsely reporting a species as present (false-positive errors). Estimating these rates allows us to quantify volunteer data quality, and may prevent the inference of erroneous trends. We use a new parameterization of a dynamic occupancy model to estimate and adjust for false-negative and false-positive errors, producing accurate estimates of occupancy. We validated this model using simulations and applied it to 12 species datasets collected from a 15-year, large-scale volunteer amphibian monitoring program. False-positive rates were low for most, but not all, species, and accounting for these errors led to quantitative differences in occupancy, although trends remained consistent even when these effects were ignored. We present a model that represents an intuitive way of quantifying the quality of volunteer monitoring datasets, and which can produce unbiased estimates of occupancy despite the presence of multiple types of observation error. Importantly, this allows the quality of volunteer monitoring data to be assessed without relying on comparisons with expert data.

DOI: https://doi.org/10.1111/csp2.54

Received: 25 September 2018 | Revised: 6 May 2019 | Accepted: 8 May 2019

DOI: 10.1111/csp2.54

CONTRIBUTED PAPER

Conservation Science and Practice
A journal of the Society for Conservation Biology
Open Access
WILEY

Quantifying data quality in a citizen science monitoring program: False negatives, false positives and occupancy trends

Sam S. Cruickshank[1] | **Christoph Bühler**[2] | **Benedikt R. Schmidt**[1,3]

[1]Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zürich, Switzerland

[2]Hintermann & Weber AG, Reinach, Switzerland

[3]info fauna karch, UniMail, Neuchâtel, Switzerland

Correspondence
Sam S. Cruickshank, Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Zürcherstrasse 111, CH-8093 Birmensdorf, Switzerland.
Email: samcruickshankresearch@gmail.com

Present address
Sam S. Cruickshank, Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Zürcherstrasse 111, CH-8093 Birmensdorf, Switzerland.

Funding information
Bundesamt für Umwelt, Grant/Award Number: 13.007PJ/M214-0617

Abstract

Data collected by volunteers are an important source of information used in species management decisions, yet concerns are often raised over the quality of such data. Two major forms of error exist in occupancy datasets; failing to observe a species when present (imperfect detection—also known as false negatives), and falsely reporting a species as present (false-positive errors). Estimating these rates allows us to quantify volunteer data quality, and may prevent the inference of erroneous trends. We use a new parameterization of a dynamic occupancy model to estimate and adjust for false-negative and false-positive errors, producing accurate estimates of occupancy. We validated this model using simulations and applied it to 12 species datasets collected from a 15-year, large-scale volunteer amphibian monitoring program. False-positive rates were low for most, but not all, species, and accounting for these errors led to quantitative differences in occupancy, although trends remained consistent even when these effects were ignored. We present a model that represents an intuitive way of quantifying the quality of volunteer monitoring datasets, and which can produce unbiased estimates of occupancy despite the presence of multiple types of observation error. Importantly, this allows the quality of volunteer monitoring data to be assessed without relying on comparisons with expert data.

KEYWORDS

citizen science, false-positive, imperfect detection, observation bias, occupancy modeling, trend, volunteer monitoring

1 | INTRODUCTION

In recognition of the potential for volunteers to allow cost-effective data collection across large spatial scales, there has been a dramatic increase in citizen-science projects over recent years (Altwegg & Nichols, 2019; Silvertown, 2009). Despite their increase in popularity, there are concerns over the utility of such volunteer-collected data (Crall et al., 2011; Lewandowski & Specht, 2015; Schmeller et al., 2009; van Strien, Pannekoek, & Gibbons, 2001), with a major issue being the perception that such data are of lower quality than that collected by professional scientists (Bird et al., 2014; Foster-Smith & Evans, 2003; Ratnieks et al., 2016). Volunteer-collected data is often the main source of information from which conservation and management decisions are made (Stem, Margoluis, Salafsky, & Brown, 2005). There is therefore a need for methods that can evaluate the reliability of datasets collected by volunteers and account for errors, thereby increasing confidence that management decisions are made using accurate and robust information.

Attempts to assess the quality of citizen-science data typically compare data collected simultaneously by volunteers and scientists (Dennis, Morgan, Brereton, Roy, & Fox, 2017; Fitzpatrick, Preisser, Ellison, & Elkinton, 2009; Forrester et al., 2015). Such approaches assume that data from expert scientists represent biological truth and any discrepancies represent errors on the behalf of volunteers. These approaches seems flawed when considering that "experts" may make errors as frequently as volunteers (Austen, Bindemann, Griffiths, & Roberts, 2016, 2018; Burgman et al., 2011), and indeed the experience of engaged amateur naturalists have long been recognized and valued as highly accurate by scientists (Altwegg & Nichols, 2019; Pocock, Roy, Preston, & Roy, 2015; Tansley, 1904). It therefore seems prudent to develop alternative methods that allow data quality to be evaluated without referring to a benchmark with unknown accuracy.

Within the context of species presence/absence monitoring, there are several ways in which observer data can contain errors. Imperfect detection, which is a consequence of false-negative errors (i.e., the failure to detect a species when present), is acknowledged as ubiquitous in monitoring data (Kellner & Swihart, 2014; Kéry & Schmidt, 2008; Preston, 1979), and statistical frameworks to estimate and account for this are well developed (Guillera-Arroita, 2017; Pollock et al., 2002; Royle & Dorazio, 2008). The relationship between the observed proportion of occupied sites, C, and the true proportion of occupied sites, N, can be considered as (Nichols, 1992):

$$C = N * p \quad (1)$$

with p representing the detection probability, which can be estimated using information from a variety of methods such as repeat surveys or distance sampling (Guillera-Arroita, 2017). A second form of bias, which has to date received much less attention, is that of false-positive errors. These occur as a result of an observer reporting a species as present when in reality the site is not occupied. In the presence of false-positives, our conceptual Equation (1) must be modified to:

$$C = (N * p) + (1 - N) * f.p \quad (2)$$

where f.p is the probability of an unoccupied site being wrongly reported as occupied. Because the model describes the two types of observation error that may occur in species presence/absence surveys, this simple equation conceptualizes a good approach to assessing the data quality of monitoring data.

Models to allow for the presence of false-positive observations were first developed by Royle and Link (2006) within the context of single-season occupancy models (MacKenzie et al., 2002). They found that simultaneously estimating both false-negative and false-positive error rates is computationally challenging, as any set of detection histories can be equally well explained by multiple sets of parameter values (Guillera-Arroita, Lahoz-Monfort, van Rooyen, Weeks, & Tingley, 2017). Royle and Link (2006) addressed this issue by forcing a constraint upon the model that the false-positive error rate must be lower than the true detection rate. Subsequent developments have developed alternative solutions to this identifiability issue by utilizing extra information to inform the detection parameters. Typically this involves jointly analyzing the dataset of interest alongside a second, independent dataset at which a subset of sites are monitored using secondary detection methods in which the probability of false-positive observations is considered impossible (Chambert, Miller, & Nichols, 2015; Miller et al., 2011, 2013). Other approaches have made use of calibration experiments to experimentally infer false-positive error rates under controlled conditions and use these to inform analysis of survey data (Chambert et al., 2015; Guillera-Arroita et al., 2017; Lahoz-Monfort, Guillera-Arroita, & Tingley, 2016; Ruiz-Gutierrez, Hooten, & Grant, 2016). These approaches have been successfully applied, however performing calibration experiments to inform false-positive error rates in survey data is often impractical (though see McClintock, Bailey, Pollock, and Semons (2010) and Ruiz-Gutierrez et al. (2016) for a successful application), and there may be situations where secondary datasets are not available. For such cases, evaluating the quality of such monitoring datasets therefore requires approaches which function without relying on restrictive constraints or extra data.

In this study, we modified the false-positive dynamic occupancy model of Miller et al. (2013) in order to evaluate the effectiveness of a volunteer amphibian monitoring program in Switzerland. We used informative priors to derive detection and occupancy rates in a Bayesian context, without the need to rely on secondary datasets or restricting parameter values. We used simulations to identify the underlying biological and detectability conditions under which our model produced reliable inference.

We applied our model to a 15-year dataset relating to 12 amphibian species at 648 sites in Switzerland, collected by approximately 250 volunteers, and we derived false-positive error rates and detection rates for each species. We compared the resulting occupancy rates and trends with both standard occupancy models (accounting for imperfect detection only), and the raw data (implicitly assuming perfect data collection), to establish how ignoring observer errors can influence the conclusions drawn from monitoring.

Our model represents a method by which the quality of occupancy datasets can be reliably assessed without the need for comparison with external datasets. Importantly, we outline the circumstances under which unmodeled errors become problematic, and demonstrate that other than in extreme situations, our model produces reliable measures of occupancy

and population trends despite the presence of these errors. We discuss that false-positive errors may arise not only from misidentification, but also through violations of model assumptions which are commonly ignored. We believe that false-positive dynamic occupancy models represent a good way of performing quality control on long-term volunteer monitoring programs, and can also be used to mitigate issues caused by the presence of transient individuals in habitat patches.

2 | METHODS

2.1 | Aargau monitoring program

Comprehensive surveying of the Swiss Canton of Aargau was carried out in the 1990s to identify and protect priority areas for amphibian conservation (Flory, 1999; Meier & Schelbert, 1999). The cantonal authorities developed an amphibian action plan in order to maintain important wetlands and improve the declining conservation status of amphibians within the region (Flory, 1999). Since 1999, approximately 900 amphibian breeding sites within the canton have been monitored by volunteers in order to evaluate the status and trends of amphibian species, and to assess the success of conservation measures that have been implemented in the area.

The monitoring program focusses specifically on 10 "core-areas" within the canton, which represent the areas surrounding the main rivers in the area (Figure 1). These core-areas each contain between 21 and 96 sites, which represent all the potential breeding sites for eight priority amphibian species that are of conservation concern within Aargau (Meier & Schelbert, 1999). The number of sites in each core area varied through time (Figure S1), as since the start of monitoring in 1999, some sites have been destroyed or otherwise converted to nonsuitable habitat. In other cases, new potential breeding habitat was created.

These core areas have been systematically surveyed since 1999 in a rotating panel design (McDonald, 2003), whereby between 1 and 5 core areas are selected each year, and all

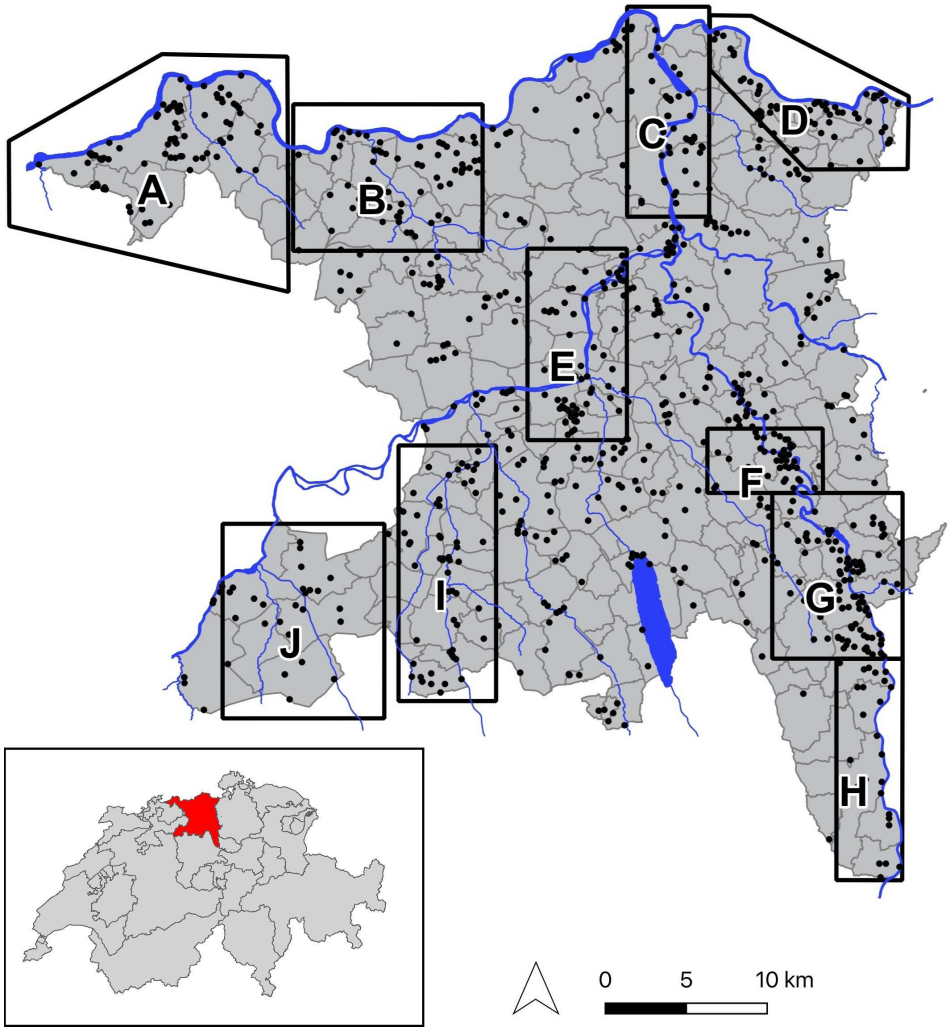

FIGURE 1 Map of the canton of Aargau, showing the 10 core-areas (boxed) that were the focus of this study, as follows: (a) Unteres Rheintal, (b) Mittleres Rheintal, (c) Unteres Aaretal, (d) Oberes Rheintal, (e) Mittleres Aaretal, (f) Unteres Reusstal, (g) Reussebene Nord, (h) Reussebene Süd, (i) Suhretal, (j) Wiggertal. Points represent amphibian breeding sites covered by the monitoring program. Inset shows the canton highlighted on a map of Switzerland. Data sources: Main Figure: Data of the canton of Aargau, Inset: Federal Office of Topography (swisstopo)

sites within those areas surveyed. Thus, each core area was comprehensively surveyed 4–5 times (mean 4.3) throughout the 15 years of monitoring data used in this study (Table S1).When surveyed, each site was visited on three occasions during predefined time windows coinciding with the amphibian breeding season. To ensure comparable survey effort between sites, the durations of surveys were allocated to be constant relative to the water surface area of the site. Volunteers were asked to record complete checklists; we therefore were able to infer non-observation of species not explicitly reported during a visit (Kéry, Gardner, & Monnerat, 2010). Further information on the survey protocol can be found in Tanadini and Schmidt (2011) and Roth, Bühler, and Amrhein (2016). In total, detection/non-detection data were collected at 648 sites for 11 species, plus one hybridogenetic species complex (*Pelophylax esculentus*: hereafter *P. escu* complex).

2.2 | General model

Our dynamic false-positive detection model is based on the Bayesian formulation of the multiple detection method of Miller et al. (2013). However, unlike the Miller model, we made use of informative priors, rather than secondary datasets, to resolve identifiability issues. We also modified our model to account for the fact that the number of sites included in our analysis changes annually (reflecting the creation or destruction of suitable habitat), and also the fact that sites were not surveyed in every year.

The initial occupancy status of site i in year $t_{initial}$ (which corresponds to the first year in which each site was first surveyed), depends on the initial occupancy probability Ψ, as

$$zi, t_{\text{initial}} \sim \text{Bernoulli}(\Psi_i) \quad (3)$$

In subsequent years, occupancy states are derived as a Markovian process (MacKenzie, Nichols, & Hines, 2003; Royle & Kéry, 2007), depending on the probability of local extinction in a site (ε) and colonization of a previously unoccupied site (γ).

$$z_{i,t+1} \sim \text{Bernoulli}\left(\text{extant}_{i,t+1} * \left(\left(z_{i,t} * (1-\varepsilon)\right) + \left(1-z_{i,t}\right) * \gamma\right)\right) \quad (4)$$

The parameter *extant* is specific to our model. It is a matrix of one's and zero's specifying, for each of the i sites, whether the site existed in each of t years; that is, the row for a site destroyed after the fourth year of monitoring would consist of four 1's followed by a string of zeros. We included this term to ensure that occupancy states were deterministically zero when we knew that a breeding site did not exist (e.g., for those years before a new pond was dug), which ensures that colonization and extinction parameters would account for this behavior.

We followed the multiple detection method of Miller et al. (2013) to link these $z_{i,t}$ underlying occupancy states to the detection/non-detection data. All occupancy models require repeat visits to sites within the primary survey periods (years), which results in observational data $y_{i,j,t}$ which represent whether or not the species was detected at site i site, during repeat visit j in year t. The data represent the underlying occupancy states as follows:

$$y_{i,j,t} \mid z_{i,t} \sim Bernoulli\left(\left(z_{i,t} * p_{11}\right) + \left(1-z_{i,t}\right) * p_{10}\right) \quad (5)$$

This ensures that at occupied sites (where $z_{i,t} = 1$), a detection will occur with probability p_{11} (the per-survey true detection probability), and at unoccupied sites a detection will occur with probability p_{10} (the per-survey false-positive error rate). However, in the absence of extra information, this model suffers from the identifiability issues identified by Royle and Link (2006), thus we added further information through the use of priors.

Rather than making the constraint of Royle and Link (2006), we instead made use of informative priors to resolve the structural non-identifiability issue and improve parameter estimation (McCarthy & Masters, 2005; Outhwaite et al., 2018; Rannala, 2002). We made use of the beta-distribution for assigning informative priors, as (i) it spans the range of possible probabilities (i.e., the interval [0,1]), and (ii) it can produce distributions ranging from flat [Beta(1,1) is equivalent to Uniform(0,1)] to highly skewed distributions [e.g., Beta(1,10), which has mode 0 and mean 0.09] (Figure S2). Having tested prior distributions ranging from Beta(1,1) to Beta(1,10), we assigned Beta (1,2) priors to three prior distributions; p_{10} (the false-positive error rate), ε (local extinction rate), and γ (the patch colonization rate). The per-survey true detection probability p_{11} was assigned a typical Uniform(0,1) prior, therefore this specification ensured that, on average $p_{11} > p_{10}$ (Figure S2). It is important to note that this prior specification does still allow for these parameters to take values up to and including 1, which is a key difference to the initial approach of Royle and Link (2006), who made the hard constraint that $p_{10} < p_{11}$. This prior choice for patch colonization and local extinction rates is based on the knowledge that patch persistence (the complement to local extinction) is greater than patch colonization rates for the species included in this study throughout Switzerland (Cruickshank, Ozgul, Zumbach, & Schmidt, 2016).

2.3 | Simulations

We ran simulations upon a simplified version of our model to test its performance under a range of scenarios, and to evaluate the influence of our choice of priors. Our simplification of the model for the simulations was that there was no addition or removal of sites over the simulated datasets (i.e., all elements of $extant_{i,t}$ had value 1).

We ran simulations to assess the performance of the model under a range of scenarios. Our primary goals were (a) to assess model performance across parameter space—particularly in relation to bias and precision (credible interval [CRI] widths) around occupancy estimates, and (b) to test the influence of different prior specifications on the false-positive error rate, colonization and local extinction rates upon the ability of the models to produce reliable inference. Additionally, we wanted to see whether assuming constant local extinction and colonization rates substantially influenced parameter estimation when the underlying data contained variation in these parameters. Finally, we wanted to assess how model performance was impacted when survey data were not available for all sites in all years—a feature seen in the rotating panel design of our case study.

We tested model performance under four different scenarios:

1. Surveys carried out in all years, no variation in colonization and local extinction rates.
2. Surveys carried out in all years, time-varying colonization and local extinction rates.
3. Scenario 1, with data for 66.6% of the years converted to NA.
4. Scenario 2, with data for 66.6% of the years converted to NA.

For each scenario, we generated 243 datasets covering a wide range of values for all key parameters (see Appendix S1 for further details on the parameter combinations analyzed under each scenario, and Appendix S2 for simulation code). To compare the effects of the information provided by priors, to each of these datasets, we applied three different version of our model; in the first, ε, γ, and p_{10} were assigned (noninformative) Beta(1,1) priors. In the second, we used slightly informative Beta(1,2) priors, and in the third version we used highly informative Beta(1,10) priors (Figure S2).

We examined our simulation results and assessed bias and uncertainty (CRI width) in key parameters (occupancy rates, p_{11}, p_{10}, colonization and local extinction), as well as the presence of any trends in occupancy bias through time. We assessed convergence using the Brooks-Gelman-Rubin statistic (Kéry & Schaub, 2011), using R-hat threshold greater than 1.1 to indicate non-convergence.

2.4 | Analysis of volunteer monitoring data

We further modified our core-model described in section 2.2 to account for characteristics specific to our study system. Surveys of the study area in the 1980s and 1990s (Flory, 1999) highlight that not all species occupied all of the 10 core-areas in our study. We therefore allowed for differences between core areas in initial occupancy rates, through the inclusion of a core-area specific dummy variable (named "historic"), which took value 1 if these earlier surveys found that the species inhabited each core area. The initial occupancy state $z_{i,initial}$ for site i, in initial survey year initial was modeled as:

$$\text{logit}\left(\text{initial occupancy}_{\text{area(i)}}\right) = \alpha_1 + \left(\alpha_2 * \text{historic}_{\text{area(i)}}\right) + \text{error}_{\text{area(i)}} \quad (6)$$

$$z_{i,\text{initial}} \sim \text{Bernoulli}\left(\text{initial.occupancy}_i * \text{extant}_{i,\text{initial}}\right) \quad (7)$$

Here, $area_i$ denotes the core area to which a given site i belongs. The coefficients α_1 and α_2 were given Uniform (−20,20) priors. We additionally included a core-area specific random effect term with standard deviation τ specified with a Gamma(1,1) prior, as:

$$\text{error}_{\text{area(i)}} \sim \text{Normal}(0, \tau) \quad (8)$$

We included temporal random effects for both detection probabilities and false-positive error rates, such that these values could vary between years. Mean rates were specified using the priors described in the simulation study, and annual realizations were drawn as normally distributed random variables around these means. Standard deviations for each rate (on the logit scale) were assigned Gamma(1,1) priors (further information provided in Appendix S3). We used the same formulation to allow colonization and local extinction rates to vary between core-areas (though remaining constant through time).

For each of the 12 species, we applied the model described above (false negative and false positive model; hereafter the FNFP-model), as well as a standard dynamic occupancy model considering only imperfect detection (false negative only model; hereafter FNO-model; MacKenzie et al., 2003). Summaries of the observer effects accounted for in each model are given in Table S2. We furthermore obtained occupancy estimates from the raw data by assuming a species was absent from a surveyed site if it went undetected in all surveys in a given year, and present if it was detected at least once. All models were run in JAGS

(Plummer, 2003) using jagsUI (Kellner, 2018) in R (R Core Team, 2017). Code for the full model can be found in Appendix S4. For each model, we ran 3 chains of 150,000 iterations with a burn-in of 10,000, thinning the remaining samples by 1 in 30. Convergence was assessed using the Brooks-Gelman-Rubin statistic (Kéry & Schaub, 2011).

For each model (FNFP and FNO) we derived the relevant detection parameters. We furthermore extracted the number of occupied sites for each core area under the different observation error scenarios from both models, and from the raw data; given that the number of sites within core area changed through time, the number of occupied sites is more informative than occupancy rates. We also derived annual trends for each core area, which we define as the absolute change in occupancy between consecutive years (i.e., $\Psi_{t-1} - \Psi_t$). For both occupancy estimates and trends, we made pairwise comparisons between models to identify differences between observation-effect scenarios. Full details of the methods used to make these comparisons are given in Appendix S5.

3 | RESULTS

3.1 | Simulations

Full results from our simulation study are presented in the form of an interactive Shiny app (https://samcruickshank.shinyapps.io/false-positive-simulations/), where the effects of various parameter combinations, prior specifications, and the presence of missing data on parameter estimation can be visualized in full. We also provide code to carry out the simulations in Appendix S2, which can be modified to test combinations of parameters of interest to readers.

Our simulations demonstrated that our model is well able to estimate occupancy and detectability parameters under a range of realistic conditions when informative priors were used. In contrast, the model did not perform well when non-informative priors were used. We found that when we used noninformative [i.e., Beta(1,1)] priors for false-positive error rates, colonization and local extinction rates, the model contained parameters which failed to converge for a high proportion of datasets (see Shiny app), and occupancy was subsequently estimated with high error and low precision. Non-convergence in these cases occurred because the structural non-identifiability means that there is equal support for several sets of possible solutions (Guillera-Arroita, 2017), and thus the Markov chains failed to mix as they got trapped on different local optima. Inference from our model with uniform priors for all parameters is therefore unreliable.

In contrast, when either Beta(1,2) or Beta(1,10) priors were instead used, model convergence was rarely an issue. Differences between the resulting estimates were very similar between models using these two prior specifications; we focus now on model results using the less informative Beta(1,2) priors.

Our results clearly show that false-positive errors can bias occupancy estimates, however only when high false-positive error rates interact with low detection probability. In simulations with a high detection probability (0.85), even high false-positive rates (0.15) do not bias occupancy estimates (Figure 2). With an intermediate detection probability of 0.55, high false-positive rates of 0.15 may induce some bias in occupancy estimates. At low levels of true-detection (0.35), the situation is exacerbated and even false-positive rates of 0.075 can lead to inaccurate occupancy estimates (Figure 2). This interaction between low detection and high false-positive rates can also lead to overestimation of colonization, decreased precision of local extinction, and underestimation of true detection (see Shiny app). However, for the range of true detection and false-positive error rates estimated for the species in our study, our model is able to accurately estimate occupancy.

In most simulations, there was no evidence of bias in the occupancy trend (see Shiny app). There were two conditions when substantially biased trends could occasionally arise (i.e., the divergence between true occupancy and estimated occupancy rate increased by more than 0.01 per year); firstly, when the highest false-positive and lowest detection rates occurred, and rarely when missing data were present and high false-positive error rates (p_{10}) combine with intermediate detection rates (p_{11}) (see Shiny app).

When we applied the model to simulated datasets containing many missing years of data, the precision of occupancy estimates decreased, and in some cases caused slight underestimation of occupancy rates when contrasted with the comparable datasets without missing data. Model convergence was occasionally problematic when models were applied to datasets where colonization and extinction rates varied, but the model did not allow for such variation; overall, 96.4% of models applied to simulation sets 1 and 3 had all parameters converge, whereas for simulation sets 2 and 4, this dropped to 89.2%.

3.2 | Case study: volunteer monitoring

There was wide variation in true detection rates between species in our volunteer dataset, and for some species detection differed substantially between the different occupancy models (Figure 3). The FNFP model always produced higher estimates of true detection probability than the FNO model. False-positive error rates were low for most species, although for three common species (*Bufo bufo*, *Ichthyosaura alpestris*, *Rana temporaria*) false-positive error rates exceeded 5%. Our simulation results suggest that the combinations of false-positive and true detection rates estimates

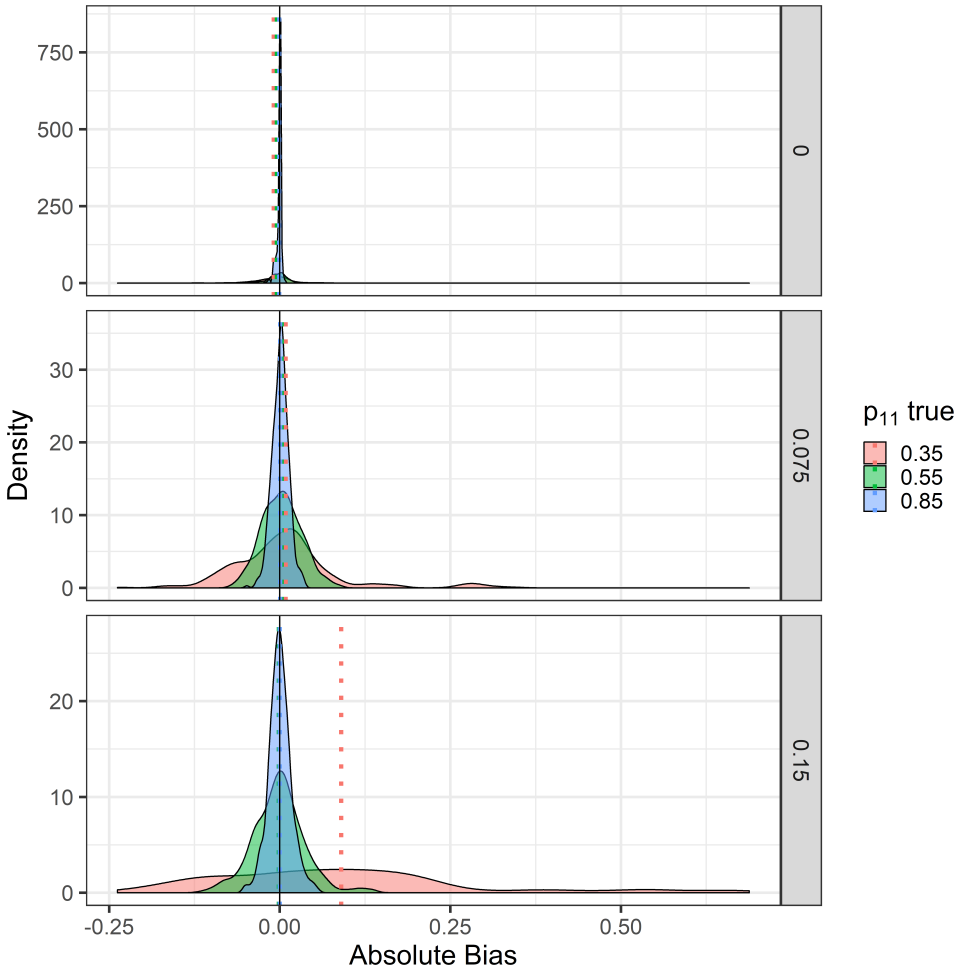

FIGURE 2 Density plot of absolute bias in occupancy estimates across all simulated datasets for simulation scenario 1. Top panel shows bias for datasets where detectability was high (0.85), middle panel shows results for simulations where detectability was 0.55, and bottom panel is for detectability of 0.35. Vertical bars represent the median bias for each group

for all species fell well within the range of parameter space for which occupancy estimates are unbiased.

The differences in the observation effects accounted for in each of the occupancy models sometimes resulted in differing occupancy estimates (Figure 4, 5; Figure S3, Table S3). For most species, a proportion of occupancy estimates differed significantly among models, although the magnitude of the differences was typically small (Table S3). However, for the four species with the highest false-positive error rates (*B. bufo, I. alpestris, R. temporaria* and the *Peloplylax esculentus* species complex), over a third of occupancy comparisons between the FNO and FNFP differed significantly from one another. When differences existed, FNO models produced higher estimates than the equivalent FNFP model (with the exception of *Lissotriton vulgaris* in the Suhretal core area; SI8). The occupancy observations from the raw data were never higher than the FNO model estimates, but were not consistently higher than those from the FNFP model (Table S3). Occupancy was estimated with higher precision in models accounting for false-positive errors: CI widths were 6.07 sites (sd: 4.19) for FNO models, and 5.66 (3.84) for FNFP models (averaged across all species). However, in line with simulation results, we observed differences among species—with lower precision for species with higher false-positive error rates.

A similar pattern was seen in our measure of population trends. For all species, there were some occasions when annual changes in occupancy rates (our measure of trend) differed among the models (Table 1). However, the magnitude of such differences was small; in every comparison, the 95% CI included zero. There was therefore no strong evidence to suggest that occupancy trends, as measured here, differ between any of the models we considered.

Therefore, with some rare exceptions when occupancy differed slightly between models (see Table S3), the occupancy rates and trends did not differ significantly between models accounting for false-positive and false-negative observation errors, models accounting only for false-negatives, and the unadjusted observational data.

4 | DISCUSSION

When volunteers are used to report presences of threatened species, there are two main errors that volunteers can make. They can either miss a species or they can report a species

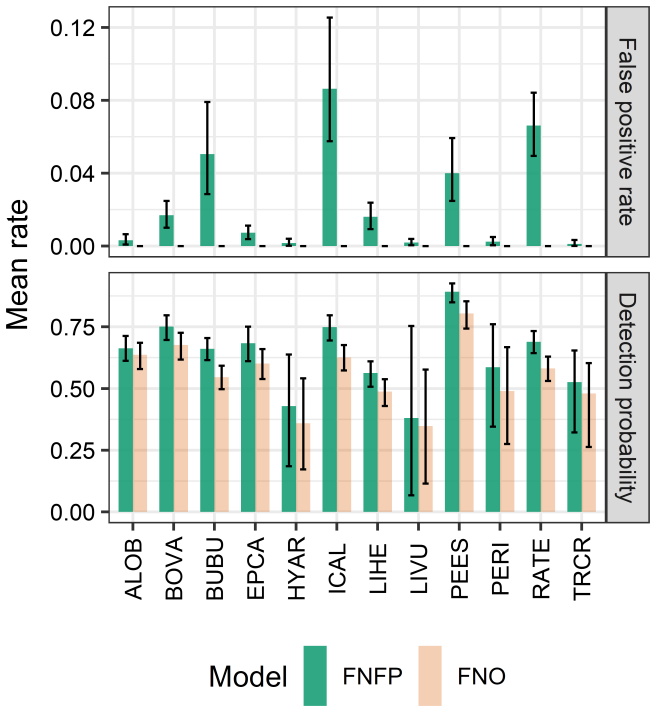

FIGURE 3 False-positive error rates, and true-detection probabilities for 12 species for our occupancy model accounting for both false-positives and false-negative errors (FNFP model). For comparison, we present the corresponding estimates of the detection probability (FNO model) from a standard occupancy model which accounts only for false-negative errors. Error bars represent 95% credible intervals. Species names are as follows: ALOB is *A. obstetricans*, BOVA is *B. variegata*, BUBU is *B. bufo*, EPCA is *E. calamita*, HYAR is *H. arborea*, ICAL is *I. alpestris*, LIHE is *L. helveticus*, LIVU is *L. vulgaris*, PEES is the *Pelophylax* species complex, PERI is *P. ridibundus*, RATE is *R. temporaria* and TRCR is *T. cristatus*

that does not occur in a particular patch. Our model parameterization provides a method by which false-negative and false-positive error rates may be estimated and accounted for in multiseason occupancy datasets, and thus the quality of volunteer-collected data may be quantitatively assessed without the need to rely on secondary data. We demonstrate that robust inference may be drawn from such datasets under a range of likely biological and sampling conditions, including when significant false-positive error rates exist, as long as detection probabilities are not also low. In our case study, we found that false-positive errors were rare for all but the most common amphibian species monitored.

One of the main benefits of false-positive occupancy models is that they allow the quality of volunteer data to be assessed directly from the dataset, rather than by requiring comparisons against other information such as expert opinion, or the need to collect additional secondary datasets. We believe that detection probabilities and false-positive error rates represent two clear metrics that can be used to evaluate the quality of volunteer datasets (see Equation (2)), and thus are a useful way to perform quality control on volunteer monitoring programs. In our dataset we found that false-positive errors were rare for most species, but that high (>5%) error rates were possible—although detection probabilities were always sufficiently high to allow accurate occupancy estimation.

Previous studies reported a tendency of volunteers to preferentially report sightings of rarer species (Lepczyk, 2005; Lewandowski & Specht, 2015; McDonough MacKenzie, Murray, Primack, & Weihrauch, 2017). There is also some evidence that false-positive error rates may be higher for rare species (Farmer, Leonard, & Horn, 2012), a phenomenon coined rare-species bias. Such a pattern is concerning, as unmodeled false-positives have more serious impacts on occupancy estimation for rare species (Miller et al., 2011). When the number of truly occupied sites is low, if even a small number of unoccupied sites are misclassified due to false-positive errors, this may drastically inflate occupancy estimates. Encouragingly, we saw no evidence for this in our study; rare species had only negligible error rates, and the only the most common species ever had high false-positive error rates. This suggests that volunteers in our study were cautious when reporting observations—if they were uncertain about the identity of a species, they appear more likely to conclude that the species was a common than a rare one. Such a tendency would inflate the false-positive rate of common species and decrease true detection probability of rarer species. We suggest that this effect is likely to hold true in volunteer systems in which the suite of potential species is relatively small, and when volunteers are aware of the goals of the program, undertake some training in identification, and receive guidelines for reporting uncertain observations. In our monitoring case-study, quality control is carried out when data is reported by volunteers; dubious reports for rare species were double-checked and excluded from the dataset if considered implausible. We note that although such a safeguard is an important and necessary feature in all monitoring programs, in our case study it was very uncommon for observations to be removed through this process, and a reanalysis of the dataset including these removed records did not change our results. Furthermore, volunteers were accepted onto the monitoring program only if they had prior experience of similar fieldwork, a willingness to self-learn the identification skills necessary, or if they had participated in a 6-day amphibian identification course.

In our analysis of amphibian populations, we identified that failing to account for imperfect detection and/or false-positive errors sometimes led to quantitative differences in occupancy rates, but that the overall trends (annual changes in occupancy) were consistent between models. This

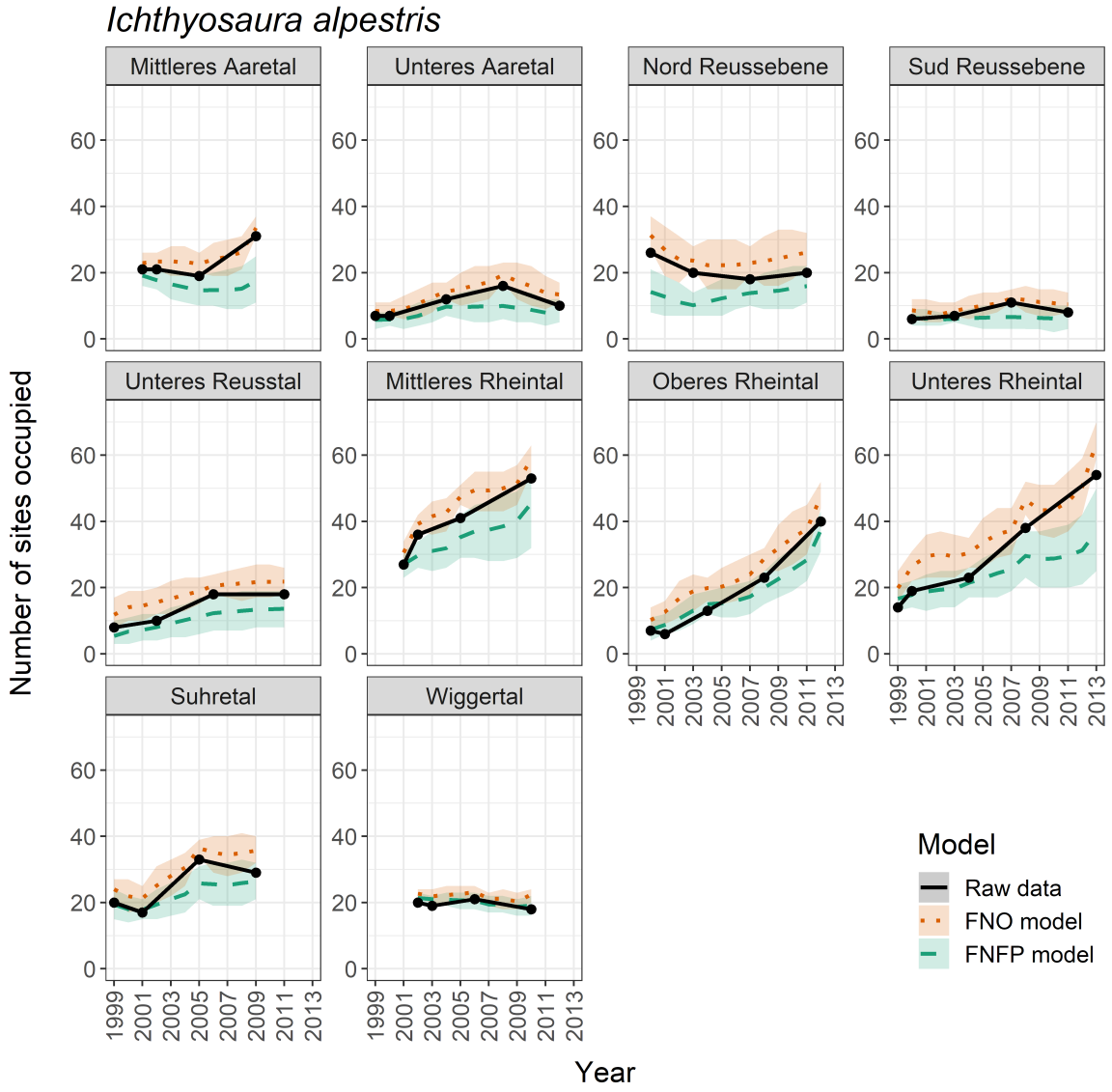

FIGURE 4 Number of occupied sites for a widespread species, the Alpine newt *Ichthyosaura alpestris*, with a high estimated detection rate (p_{11}) and false-positive rate (p_{10}) over 10 core areas (panels). Ribbons for the FNO and FNFP represent the 95% credible interval. The line for raw data represents the number of sites at which observations were made in the dataset (i.e., assuming no false-positives and perfect detection)

suggests that the benefits of accounting for such errors may depend strongly on the goals of the monitoring program. For a conservation-manager trying to identify all sites that require conservation management, considering false-positives in monitoring data would help focus management at the appropriate locations. However, if the goal is long-term monitoring of population trends, our results suggest that the benefits of accounting for false-positive errors were slight. Consequently, program managers may decide that the results from standard occupancy modeling may be a sufficiently satisfactory method to achieve this goal.

In programs designed to monitor population status and dynamics, we may consider two sources of possible false-positive errors. The first, which we can consider as sampling false-positives, arise when surveyors report seeing a species when it is not in fact present at (e.g., through misidentification or wrong field notes). However, we may consider a second source of false-positives, termed ecological false-positives by Berigan, Jones, Whitmore, Gutiérrez, and Peery (2019), when detections are also made of individuals which are temporarily making use of a site outside of their home range—for example, when individuals are dispersing (Sutherland, Elston, & Lambin, 2013), or foraging outside of their normal home range (Berigan et al., 2019). The presence of such observations will mean that traditional analysis provides inference on site-use, rather than occupancy patterns in the strict sense of the term (Altwegg & Nichols, 2019). Treating such records of transients as evidence for the presence of an established populations therefore results in biased inference (Altwegg & Nichols, 2019; Berigan

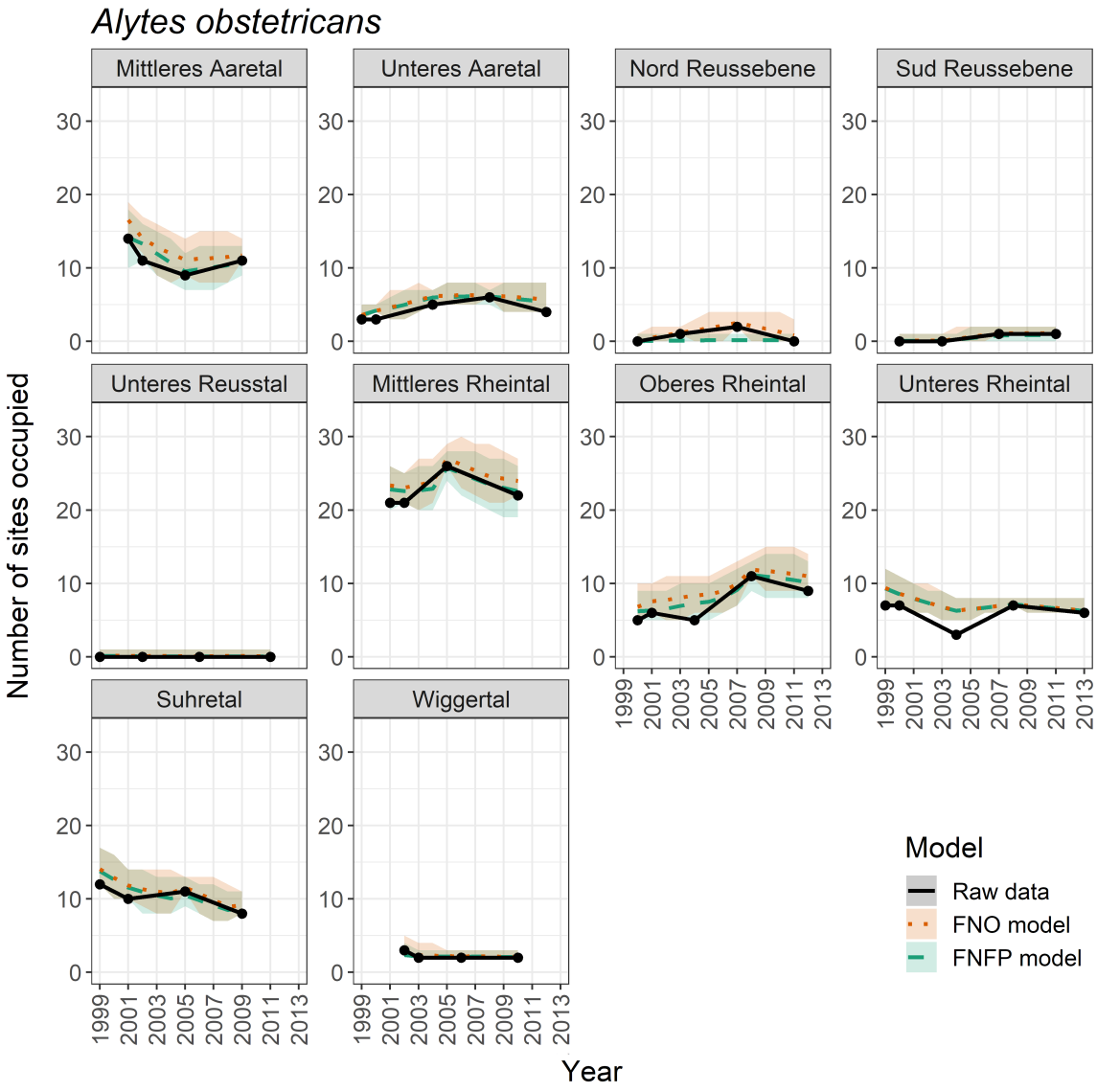

FIGURE 5 Number of occupied sites for a rare species, the common Midwife toad *Alytes obstetricans* over 10 core areas (panels). Ribbons for the FNO and FNFP models represent the 95% credible interval. The line for raw data represents the number of sites at which observations were made in the dataset (i.e., assuming no false-positives and perfect detection)

et al., 2019; Latif, Ellis, & Amundson, 2016), and leads to overestimation of occupancy (Rota, Fletcher, Dorazio, & Betts, 2009). Consider detection histories from 3 surveys at 3 sites:

Site 1 : 101 Site 2 : 000 Site 3 : 000

Here a standard occupancy model, which calculates detection probabilities from those sites where detections were made, would produce a detection estimate of 0.66 (i.e., $\frac{2}{3}$), and site 1 would be classified as occupied. The probability of obtaining 3 nondetections if the site is truly occupied is only 3.6% (i.e., $(1-0.66)^3$; McArdle, 1990, Wintle, Walshe, Parris, & McCarthy, 2012), thus the model is unlikely to conclude that sites 2 or 3 are occupied. In contrast, by introducing a detection of one transient into the detection history of site 2, we can see that there are 2 routes by which transients lead to unoccupied sites being wrongly classified as occupied (under the assumption, that a site should only be classified as "occupied" if there is a population; transient individuals are not sufficient):

Site 1 : 101 Site 2 : 010 Site 3 : 000

First, it is clear that this detection of a transient will lead to 2 out of the 3 sites being classified as occupied by a standard model. However, the information used to estimate detection probabilities changes; now 6 surveys are used to calculate detection probability, and the estimated probability declines from 0.66 to 0.50 ($\frac{3}{6}$). Consequently, we also have

TABLE 1 The proportion of occupancy trends that are significantly different from 0, and the overall mean difference and 95% credible intervals of annual trends for pairwise model comparisons for 12 species

	Proportion significant			Mean difference in occupancy trend (CI)		
Species name	**FNO-FNFP**	**FNO-raw**	**FNFP-raw**	**FNO-FNFP**	**FNO-raw**	**FNFP-raw**
Alytes obstetricans	0.00	0.19	0.29	0.01 (−3.00:2.00)	−0.14 (−5.25:2.00)	−0.15 (−5.25:1.67)
Bombina variegata	0.00	0.28	0.31	0.04 (−4.00:3.00)	0.05 (−5.50:5.00)	0.02 (−5.00:4.25)
Bufo bufo	0.00	0.14	0.27	−0.13 (−7.00:5.00)	0.62 (−7.00:6.00)	0.74 (−5.50:5.00)
Epidalea calamita	0.00	0.11	0.24	−0.07 (−3.00:2.00)	−0.04 (−3.25:2.33)	0.03 (−2.67:2.00)
Hyla arborea	0.00	0.13	0.27	0.08 (−2.00:2.00)	0.46 (−2.67:3.00)	0.37 (−2.50:2.50)
Icthyosaura alpestris	0.00	0.20	0.21	0.74 (−7.00:6.00)	1.12 (−7.67:8.00)	0.38 (−6.33:5.5)
Lissotriton helveticus	0.00	0.13	0.19	0.16 (−4.00:3.00)	0.11 (−5.33:3.50)	−0.05 (−5.00:3.00)
Lissotriton vulgaris	0.00	0.09	0.15	−0.08 (−3.00:2.00)	−0.06 (−3.00:1.80)	0.02 (−2.50:1.20)
Pelophylax esculentus species complex	0.01	0.18	0.19	0.24 (−5.00:4.00)	0.76 (−10.33:5.50)	0.53 (−11.33:4.75)
Pelophylax ridibundus	0.00	0.06	0.07	−0.06 (−4.00:2.00)	−0.10 (−3.80:2.00)	−0.04 (−3.00:2.00)
Rana temporaria	0.01	0.18	0.23	0.13 (−8.00:6.00)	0.84 (−11.00:8.00)	0.70 (−8.33:6.25)
Triturus cristatus	0.00	0.26	0.28	0.06 (−2.00:1.00)	0.14 (−5.00:1.67)	0.09 (−5.00:1.67)
	0.00	*0.16*	*0.22*	*0.09*	*0.31*	*0.22*

Note. Raw denotes the raw data (which implicitly assumes perfect detection and no false-positives). Italicised row denotes column means. Bold values denote that 95% credible intervals do not overlap with 0.

less confidence that site 3 is truly unoccupied, as the probability of not detecting it even if it is there also increases to 12.5% (i.e., $(1-0.50)^3$). Over large numbers of sites, this effect can lead to many sites being falsely classified as occupied, even if there are no detections made at the site. This pattern can be seen in our results on tree-frogs (*H. arborea*); this species is typically easily detectable from calls (Cruickshank et al., 2016; Pellet & Schmidt, 2005), yet our standard occupancy model (FNO model) produced detection estimates substantially lower than is typical for this species. Evidence exists that individual males frequently travel substantial distances between ponds within a breeding season in our study area (Angelone, Kienast, & Holderegger, 2011; see also Schmidt & Pellet, 2005), and the species has recently been colonizing the core-area Reussebene Nord (C. Bühler, unpublished data). We believe that occasional detections of these dispersing individuals is the cause of our unexpectedly-low detection probabilities. The best solution to the issue of transient individuals (ecological false positives) is the careful definition of sampling units when designing a study (Altwegg & Nichols, 2019; Petranka, Smith, & Scott, 2004). The usual approach "pond = population" may be wrong because within-season movements between sites appear to be common in amphibians (Kopecky, Vojar, & Denoel, 2010; Petranka et al., 2004; Schmidt, 2005; Tournier, Besnard, Tournier, & Cayuela, 2017). In situations where ecological false-positives cannot be prevented during the planning stage of monitoring, models such as ours—which can account for both forms of false-positive observations—are recommended in order to improve inference of occupancy dynamics and trends (Altwegg & Nichols, 2019; Sutherland et al., 2013).

Allowing for false-positives in occupancy modeling requires the incorporation of more information than is present in simple detection/non-detection datasets that are used in standard occupancy models. Existing approaches deal with this issue either by using calibration studies (e.g., with DNA analysis; Lahoz-Monfort et al., 2016), by making strict assumptions regarding parameter values (Royle & Link, 2006), or by using auxiliary datasets to obtain extra information (Chambert et al., 2015; Miller et al., 2011). In many cases such extra data are not available, and we must turn to alternative approaches to provide extra information to the model. Our solution to this problem was to incorporate informative priors which are based on realistic assumptions. As the simulations show, the use of informative priors helped to overcome identifiability issues (Miller et al., 2011; Royle & Link, 2006), while allowing detectability rates to vary among years. This is considered essential in dynamic models to avoid bias (Miller et al., 2015). Our approach produces accurate estimates of false-positive and occupancy rates, as well as detection probabilities. We note that this approach of adding information in the form of priors produced comparable estimates to approaches that make use of extra information. We therefore believe that making use of informative priors, when justified by biological characteristics of the system being monitored, provides a good alternative way for assessing the quality of data when

supplementary datasets are not available. Importantly, our simulation results show that strongly informative priors (which may have a strong influence on the posterior estimates from the model) were not necessary; weakly informative priors are sufficient to avoid the identifiability issues that present challenges when accounting for false-positives (Guillera-Arroita et al., 2017; Miller et al., 2011; Royle & Link, 2006).

Species monitoring studies have adapted to an understanding of the effects of imperfect detection, and are increasingly adjusting protocols to collect the necessary information to account for this issue (Kellner & Swihart, 2014). A growing recognition that false-positive errors may also be widespread in monitoring data, and the associated modeling developments that allow for these errors to be accounted for, is welcome. We believe that a secondary benefit of such models is that it can act to increase confidence in the data of volunteers—who provide massive amounts of monitoring data at low cost. By quantifying error rates, program managers can assess how accurate the data from monitoring programs are, and this can be used to counter concerns as to data quality, or to make adjustments to the protocols used in monitoring if necessary. The consequences of false-positive errors are most extreme for rare species (Miller et al., 2011), which may also be hard to detect due to low abundances (McCarthy et al., 2013; Tanadini & Schmidt, 2011). Our results suggest that false-positive errors only cause serious bias in combination with low detection probabilities. This suggests that modifying survey protocols to maximize species detection probability in a survey is likely to be the best way of ensuring accurate estimation of occupancy rates. However, modifications to minimize false-positive errors, such as informing volunteers as to how to report uncertain observations, requesting photographs of rare species reports, or carrying out other data-quality assessments, may be more easily achieved; such efforts have previously been advocated (Lovell, Hamer, Slotow, & Herbert, 2009; Miller et al., 2015).

We have presented a general model to objectively evaluate the quality of monitoring data without the need for comparisons with expert data which may itself be flawed (Austen et al., 2016; Culverhouse, Williams, Reguera, Herry, & Gonzalez Gil, 2003). We believe this will be particularly useful for program managers who wish to ensure that their survey protocols are robust and who wish to ensure that the resulting data are of high quality. In particular, we propose that the model be used if there is any suspicion that false detections of a species may be common (for example if species identification is particularly difficult). Furthermore, when it is believed that individuals may make occasional use of adjacent survey units, and this cannot be prevented during survey design, results from our model will better reflect true occupancy patterns, rather than patterns in habitat use, which would be produced by more traditional occupancy models. We therefore think that scientists and managers can use the model to fully access the information contained in data collected by citizen scientists.

ACKNOWLEDGMENTS

This work was supported by the Swiss Federal Office of the Environment (Forschungsvertrag 13.007PJ/M214-0617). We thank the Abteilung Landschaft und Gewässer of the canton of Aargau for providing permission to use the data from their amphibian monitoring program. We thank Arpat Ozgul and the reviewers for useful comments on the manuscript, and we are grateful to the many volunteers who did the field work for so many years and collected the data analyzed in this study.

DATA ACCESSIBILITY

The monitoring data used in this study is available from Figshare (doi:10.6084/m9.figshare.8075186) (Cruickshank, Bühler, & Schmidt, 2019).

AUTHOR CONTRIBUTIONS

B.R.S. and S.S.C. conceived the study design. C.B. manages the data collection for our case study, and provided data and insights into the study system. S.S.C. developed the model and carried out all analysis. All authors were involved in interpreting the results. S.S.C and B.R.S wrote the manuscript with contributions from C.B.

ETHICS STATEMENT

The authors are not aware of any ethical issues regarding this work. Data collection was not carried out specifically for this project, but was carried out in compliance with all applicable laws.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Sam S. Cruickshank https://orcid.org/0000-0003-0183-5352
Benedikt R. Schmidt https://orcid.org/0000-0002-4023-1001

REFERENCES

Altwegg, R., & Nichols, J. D. (2019). Occupancy models for citizen-science data. *Methods in Ecology and Evolution*, *10*, 8–21.

Angelone, S., Kienast, F., & Holderegger, R. (2011). Where movement happens: Scale-dependent landscape effects on genetic differentiation in the European tree frog. *Ecography*, *34*, 714–722.

Austen, G. E., Bindemann, M., Griffiths, R. A., & Roberts, D. L. (2016). Species identification by experts and non-experts: Comparing images from field guides. *Scientific Reports*, *6*, 33634.

Austen, G. E., Bindemann, M., Griffiths, R. A., & Roberts, D. L. (2018). Species identification by conservation practitioners using online images: Accuracy and agreement between experts. *PeerJ*, *6*, e4157.

Berigan, W. J., Jones, G. M., Whitmore, S. A., Gutiérrez, R. J., & Peery, M. Z. (2019). Cryptic wide-ranging movements lead to upwardly biased occupancy in a territorial species. *Journal of Applied Ecology*, *56*, 470–480.

Bird, T. J., Bates, A. E., Lefcheck, J. S., Hill, N. A., Thomson, R. J., Edgar, G. J., … Frusher, S. (2014). Statistical solutions for error and bias in global citizen science datasets. *Biological Conservation*, *173*, 144–154.

Burgman, M. A., McBride, M., Ashton, R., Speirs-Bridge, A., Flander, L., Wintle, B., … Twardy, C. (2011). Expert status and performance. *PLoS One*, *6*, 1–7.

Chambert, T., Miller, D. A. W., & Nichols, J. D. (2015). Modeling false positive detections in species occurrence data under different study designs. *Ecology*, *96*, 332–339.

Crall, A. W., Newman, G. J., Stohlgren, T. J., Holfelder, K. A., Graham, J., & Waller, D. M. (2011). Assessing citizen science data quality: An invasive species case study. *Conservation Letters*, *4*, 433–442.

Cruickshank, S. S., Bühler, C., & Schmidt, B. R. (2019). Long term amphibian monitoring data from canton of Aargau. *Figshare*. https://doi.org/10.6084/m9.figshare.8075186

Cruickshank, S. S., Ozgul, A., Zumbach, S., & Schmidt, B. R. (2016). Quantifying population declines based on presence-only records for red-list assessments. *Conservation Biology*, *30*, 1112–1121.

Culverhouse, P. F., Williams, R., Reguera, B., Herry, V., & Gonzalez Gil, S. (2003). Do experts make mistakes? A comparison of human and machine labelling of dinoflagellates. *Marine Ecology Progress Series*, *247*, 17–25.

Dennis, E. B., Morgan, B. J. T., Brereton, T. M., Roy, D. B., & Fox, R. (2017). Using citizen science butterfly counts to predict species population trends. *Conservation Biology*, *31*, 1350–1361.

Farmer, R. G., Leonard, M. L., & Horn, A. G. (2012). Observer effects and avian-call-count survey quality: Rare-species biases and over-confidence. *The Auk*, *129*, 76–86.

Fitzpatrick, M. C., Preisser, E. L., Ellison, A. M., & Elkinton, J. S. (2009). Observer bias and the detection of low-density populations. *Ecological Applications*, *19*, 1673–1679.

Flory, C. (1999). Amphibien im Kanton Aargau. *Mitteilungen der aargauischen Naturforschenden Gesellschaft*, *35*, 5–39.

Forrester, G., Baily, P., Conetta, D., Forrester, L., Kintzing, E., & Jarecki, L. (2015). Comparing monitoring data collected by volunteers and professionals shows that citizen scientists can detect long-term change on coral reefs. *Journal for Nature Conservation*, *24*, 1–9.

Foster-Smith, J., & Evans, S. M. (2003). The value of marine ecological data collected by volunteers. *Biological Conservation*, *113*, 199–213.

Guillera-Arroita, G. (2017). Modelling of species distributions, range dynamics and communities under imperfect detection: Advances, challenges and opportunities. *Ecography*, *40*, 281–295.

Guillera-Arroita, G., Lahoz-Monfort, J., van Rooyen, A., Weeks, A., & Tingley, R. (2017). Dealing with false-positive and false-negative errors about species occurrence at multiple levels. *Methods in Ecology and Evolution*, *38*, 42–49.

Kellner, K. F. (2018). jagsUI: A wrapper around 'rjags' to streamline 'JAGS' analyses. R package version 1.5.0. Retrieved from https://CRAN.R-project.org/package=jagsUI.

Kellner, K. F., & Swihart, R. K. (2014). Accounting for imperfect detection in ecology: A quantitative review. *PLoS One*, *9*, e111436.

Kéry, M., Gardner, B., & Monnerat, C. (2010). Predicting species distributions from checklist data using site-occupancy models. *Journal of Biogeography*, *37*, 1851–1862.

Kéry, M., & Schaub, M. (2011). *Bayesian population analysis using WinBUGS*. Waltham, MA: Academic Press.

Kéry, M., & Schmidt, B. R. (2008). Imperfect detection and its consequences for monitoring for conservation. *Community Ecology*, *9*, 207–216.

Kopecky, O., Vojar, J., & Denoel, M. (2010). Movements of Alpine newts (*Mesotriton alpestris*) between small aquatic habitats (ruts) during the breeding season. *Amphibia-Reptilia*, *31*, 109–116.

Lahoz-Monfort, J., Guillera-Arroita, G., & Tingley, R. (2016). Statistical approaches to account for false-positive errors in environmental DNA samples. *Molecular Ecology Resources*, *16*, 673–685.

Latif, Q. S., Ellis, M. M., & Amundson, C. L. (2016). A broader definition of occupancy: Comment on Hayes and Monfils. *The Journal of Wildlife Management*, *80*, 192–194.

Lepczyk, C. A. (2005). Integrating published data and citizen science to describe bird diversity across a landscape. *Journal of Applied Ecology*, *42*, 672–677.

Lewandowski, E., & Specht, H. (2015). Influence of volunteer and project characteristics on data quality of biological surveys. *Conservation Biology*, *29*, 713–723.

Lovell, S., Hamer, M., Slotow, R., & Herbert, D. (2009). An assessment of the use of volunteers for terrestrial invertebrate biodiversity surveys. *Biodiversity and Conservation*, *18*, 3295–3307.

MacKenzie, D. I., Nichols, J. D., & Hines, J. (2003). Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology*, *84*, 2200–2207.

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A., & Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, *83*, 2248–2255.

McArdle. (1990). When are rare species not there? *Oikos*, *57*, 276–277.

McCarthy, M. A., & Masters, P. (2005). Profiting from prior information in Bayesian analyses of ecological data. *Journal of Applied Ecology*, *42*, 1012–1019.

McCarthy, M. A., Moore, J. L., Morris, W. K., Parris, K. M., Garrard, G. E., Vesk, P. A., … Yue, B. (2013). The influence of abundance on detectability. *Oikos*, *122*, 717–726.

McClintock, B., Bailey, L., Pollock, K., & Semons, T. (2010). Unmodeled observation error induces bias when inferring patterns and dynamics of species occurrence via aural detections. *Ecology*, *91*, 2446–2454.

McDonald, T. L. (2003). Review of environmental monitoring methods: Survey designs. *Environmental Monitoring and Assessment*, *85*, 277–292.

McDonough MacKenzie, C., Murray, G., Primack, R., & Weihrauch, D. (2017). Lessons from citizen science: Assessing volunteer-collected plant phenology data with Mountain Watch. *Biological Conservation*, *208*, 121–126.

Meier, C., & Schelbert, B. (1999). Amphibienschutzkonzept Kanton Aargau. *Mitteilungen der aargauischen Naturforschenden Gesellschaft*, *35*, 41–70.

Miller, D. A. W., Bailey, L. L., Campbell Grant, E. H., McClintock, B. T., Weir, L. A., & Simons, T. R. (2015). Performance of species occurrence estimators when basic assumptions are not met: A test using field data where true occupancy status is known. *Methods in Ecology and Evolution*, *6*, 557–565.

Miller, D. A. W., Nichols, J. D., Gude, J. A., Rich, L. N., Podruzny, K. M., Hines, J. E., & Mitchell, M. S. (2013). Determining occurrence dynamics when false positives occur: Estimating the range dynamics of wolves from public survey data. *PLoS One*, *8*, e65808.

Miller, D. A. W., Nichols, J. D., McClintock, B. T., Grant, E. H. C., Bailey, L. L., & Weir, L. A. (2011). Improving occupancy estimation when two types of observational error occur: Non-detection and species misidentification. *Ecology*, *92*, 1422–1428.

Nichols, J. D. (1992). Capture- recapture models: Using marked animals to study population dynamics. *Bioscience*, *42*, 94–102.

Outhwaite, C. L., Chandler, R. E., Powney, G. D., Collen, B., Gregory, R. D., & Isaac, N. J. B. (2018). Prior specification in Bayesian occupancy modelling improves analysis of species occurrence data. *Ecological Indicators*, *93*, 333–343.

Pellet, J., & Schmidt, B. R. (2005). Monitoring distributions using call surveys: Estimating site occupancy, detection probabilities and inferring absence. *Biological Conservation*, *123*, 27–35.

Petranka, J. W., Smith, C. K., & Scott, A. F. (2004). Identifying the minimal demographic unit for monitoring pond-breeding amphibians. *Ecological Applications*, *14*, 1065–1078.

Plummer, M. (2003). *JAGS: A program for analysis of bayesian graphical models using Gibbs sampling*. Proceedings of the 3rd international workshop on distributed statistical computing 124.

Pocock, M. J. O., Roy, H. E., Preston, C. D., & Roy, D. B. (2015). The biological records Centre: A pioneer of citizen science. *Biological Journal of the Linnean Society*, *115*, 475–493.

Pollock, K. H., Nichols, J. D., Simons, T. R., Farnsworth, G. L., Bailey, L. L., & Sauer, J. R. (2002). Large scale wildlife monitoring studies: Statistical methods for design and analysis. *Environmetrics*, *13*, 105–119.

Preston, F. W. (1979). The invisible birds. *Ecology*, *60*, 451–454.

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rannala, B. (2002). Identifiability of parameters in MCMC Bayesian inference of phylogeny. *Systematic Biology*, *51*, 754–760.

Ratnieks, F. L. W., Schrell, F., Sheppard, R. C., Brown, E., Bristow, O. E., & Garbuzov, M. (2016). Data reliability in citizen science: Learning curve and the effects of training method, volunteer background and experience on identification accuracy of insects visiting ivy flowers. *Methods in Ecology and Evolution*, *7*, 1226–1235.

Rota, C. T., Fletcher, R. J., Dorazio, R. M., & Betts, M. G. (2009). Occupancy estimation and the closure assumption. *Journal of Applied Ecology*, *46*, 1173–1181.

Roth, T., Bühler, C., & Amrhein, V. (2016). Estimating effects of species interactions on populations of endangered species. *The American Naturalist*, *187*, 457–467.

Royle, J. A., & Dorazio, R. M. (2008). *Hierarchical modeling and inference in ecology: The analysis of data from populations, metapopulations and communities*. Amsterdam, Netherlands: Academic Press.

Royle, J. A., & Kéry, M. (2007). A Bayesian state-space formulation of dynamic occupancy models. *Ecology*, *88*, 1813–1823.

Royle, J. A., & Link, W. (2006). Generalized site occupancy models allowing for false positive and false negative errors. *Ecology*, *87*, 835–841.

Ruiz-Gutierrez, V., Hooten, M. B., & Grant, E. H. C. (2016). Uncertainty in biological monitoring: A framework for data collection and analysis to account for multiple sources of sampling bias. *Methods in Ecology and Evolution*, *7*, 900–909.

Schmeller, D. S., Henry, P.-Y., Julliard, R., Gruber, B., Clobert, J., Dziock, F., … Henle, K. (2009). Advantages of volunteer-based biodiversity monitoring in Europe. *Conservation Biology*, *23*, 307–316.

Schmidt, B. R. (2005). Monitoring the distribution of pond-breeding amphibians when species are detected imperfectly. *Aquatic Conservation*, *15*, 681–692.

Schmidt, B. R., & Pellet, J. (2005). Relative importance of population processes and habitat characteristics in determining site occupancy of two anurans. *Journal of Wildlife Management*, *69*, 884–893.

Silvertown, J. (2009). A new dawn for citizen science. *Trends in Ecology & Evolution*, *24*, 467–471.

Stem, C., Margoluis, R., Salafsky, N., & Brown, M. (2005). Monitoring and evaluation in conservation: A review of trends and appraoches. *Conservation Biology*, *19*, 295–309.

Sutherland, C., Elston, D. A., & Lambin, X. (2013). Accounting for false positive detection error induced by transient individuals. *Wildlife Research*, *40*, 490–498.

Tanadini, L. G., & Schmidt, B. R. (2011). Population size influences amphibian detection probability: Implications for biodiversity monitoring programs. *PLoS One*, *6*, e28244.

Tansley, A. (1904). The problems of ecology. *New Phytologist*, *3*, 191–200.

Tournier, E., Besnard, A., Tournier, V., & Cayuela, H. (2017). Manipulating waterbody hydroperiod affects movement behaviour and occupancy dynamics in an amphibian. *Freshwater Biology*, *62*, 1768–1782.

van Strien, A. J., Pannekoek, J., & Gibbons, D. W. (2001). Indexing European bird population trends using results of national monitoring schemes: A trial of a new method. *Bird Study*, *48*, 200–213.

Wintle, B. A., Walshe, T. V., Parris, K. M., & McCarthy, M. A. (2012). Designing occupancy surveys and interpreting non-detection when observations are imperfect. *Diversity and Distributions*, *18*, 417–424.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Cruickshank SS, Bühler C, Schmidt BR. Quantifying data quality in a citizen science monitoring program: False negatives, false positives and occupancy trends. *Conservation Science and Practice*. 2019;1:e54. https://doi.org/10.1111/csp2.54